

# 基于金融指标对中国 GDP 的混频预测分析

郑挺国 尚玉皇

(厦门大学王亚南经济研究院, 福建厦门 361105)

**摘要:** 本文在实时数据基础上选取金融变量作为预测因子并通过混频数据抽样(MIDAS)模型对 GDP 增长率进行短期预测。结果表明: 短期预测时 MIDAS 模型预测效果甚佳而且嵌入自回归项的 MIDAS 模型明显降低预测误差; 数据修正对 MIDAS 模型的预测精度有负面影响; 货币供应量等预测因子在包含自回归项 MIDAS 模型中预测精度较高, 表明投资和出口依旧是拉动我国经济增长的重要因素; SPA 检验及组合 MIDAS 模型的较好预测精度说明组合 MIDAS 模型预测能力占优。

**关键词:** 金融变量; GDP 增长率; MIDAS 模型; 混频预测; 数据修正

**JEL 分类号:** C32; E32; E52 **文献标识码:** A **文章编号:**

## 一、引言

经济增长是货币政策的最终目标之一, 而国内生产总值(GDP)作为重要的宏观经济变量在刻画经济增长运行态势中发挥着不可替代的作用, 但无论是政府部门还是社会公众都更加关注 GDP 的未来走势, 以期更好地制定宏观政策、规划消费决策(Heij 等人, 2011)。在恰当时刻做出合理决定需要实际经济行为人准确及时地预测宏观经济总量, 因此如何“又好又快”地预测宏观经济总量显得尤为关键。在预测中如何能够兼得“精确性”与“时效性”的问题引起了许多学者的广泛关注。根据 Koenig 等人(2003), Clements 和 Galv ão (2008, 2009), 以及 Ghysels 和 Wright (2009) 等学者的研究, 我们可以得知数据修正、预测因子和数据频率是影响预测精确性和时效性的三个重要方面。

首先, 许多宏观经济变量如 GDP 会面临数据修正问题。数据修正将最初实时数据转化为最终修正数据, 但同时也会影响模型的预测效果。为了避免数据修正对模型预测的影响, 越来越多的研究基于实时数据。Koenig 等人(2003)则强调使用实时数据进行估计和预测的效果好于最终数据, 并且最好是选择官方初次公布的数据作为实时数据。因此, 基于实时数据的预测可以减少数据修正对预测精确度及时效性的负面影响。

其次, 预测因子对模型的预测效果也会产生重要影响。关于宏观经济总量的预测, 既可以选择固定资产投资等一致指标, 也可以采用货币供应量、消费者信心指数等领先指标。由于对宏观经济总量的预测模型一般为自回归分布滞后模型(Galv ão, 2010), 因此使用领先因子特别是金融变量对宏观经济总量进行预测受到越来越多的关注, 如 Stock 和 Watson (2003), Ghysels 和 Wright (2009)。特别地, Stock 和 Watson 指出金融变量对于宏观经济变量的预测能力具有统计上的显著性。

金融变量可以成为预测宏观经济变量重要的领先因子, 主要原因是其包含了未来经济活动的预期信息。一方面, 股票等金融资产的价格等价于企业未来收益的期望贴现值, 而企业未来收益必然受到未来经济活动的影响。金融变量的前瞻性表明债券和股票收益率可以作为性质良好的预测因子(Stock 和 Watson, 2003)。消费与市场波动率间的关系也非常密切, Bansal and Yaron (2004) 基于

**收稿日期:**

**作者简介:** 郑挺国, 经济学博士, 厦门大学王亚南经济研究院副教授, Email: zhengt@163.com;

尚玉皇, 经济学硕士, 厦门大学王亚南经济研究院博士生, Email: jnuyuhuang@163.com.

本研究得到国家自然科学基金项目“货币政策规则非线性的理论模型与计量研究”(71001087)、“状态空间混频模型及其在宏观经济中的应用”(71371160)、福建省教育厅项目“我国经济波动与经济政策的 DSGE 计量建模及应用研究”(2009100074)以及国家留学基金委公派访问学者(含博士后研究)项目(201208350111)资助。作者衷心感谢匿名审稿人的有益评论和宝贵建议, 当然文责自负。

C-CAPM模型提出使用长期风险来解释超额波动率之谜。另一方面, Romer (1990)提出的“不确定性假说”观点认为股票市场波动会增加未来经济活动的不确定性,从而导致居民的消费支出减少。无论是股票的持有者还是非持有者都会将股票波动率作为未来经济活动的领先因子。另外,金融变量还表现出较强的可得性,该特征对提高预测的时效性有很大帮助。

最后,增加预测因子的数据频率丰富数据信息也是改善预测效果的关键环节。高频数据中包含的更多有效信息在提高预测精确度的同时,对预测的时效性也有着立竿见影的作用。由于宏观经济总量是季度数据,提高预测因子的数据频率会面临混频数据问题。如何处理混频数据是本文需要重点关注的核心问题之一。起初学者的解决思路是将混频数据同频化,如同期加总方法,该方法可能会造成某些重要的信息损失,在一定条件下会产生估计的有偏性和不一致性(Andreou 等人, 2010);另外还可以采取插值的方法,但是人们很难评价哪一种插值方法是合适的,而且产生的高频数据会受到测量误差的干扰(Guerin 等人, 2011)。

基于混频数据直接构建计量经济模型是另一种解决思路,许多文献通过状态空间模型(因子模型)讨论混频数据建模问题,如Camacho和 Perez-Quiros (2010),郑挺国和王霞(2013)等。基于Kalman滤波方法进行估计的因子模型还可以很好地解决数据缺失问题,但是状态空间模型有时需要估计很多参数,产生计算成本。近年来, Ghysels 等人(2004、2007)提出的混频数据抽样模型(MIDAS)<sup>①</sup>受到了大量关注。与因子模型相比混频数据抽样模型的设定相对简洁,其重要特征是通过加权函数以简洁的方式将高频数据解释变量嵌入到回归模型中。MIDAS模型的主要优点是当存在混频数据样本时,其可以更多地考虑解释变量所包含的信息量并提高估计的有效性,消除估计的有偏性,而且它能够利用最新公布的高频数据提高预测的时效性和精确性。

因此,基于金融变量的良好性质和 MIDAS 模型在处理混频数据时的优越性,许多国外学者将该方法应用于宏观经济变量实时预测的实证研究中,如 Clements 和 Galv ão (2008; 2009) 与 Ghysels 和 Wright (2009)。其中, Clements 和 Galv ão (2009)使用 MIDAS 模型对美国的经济增长进行实时预测,结果表明通过使用实时数据领先因子具有显著的预测能力。

关于我国宏观经济总量的预测,国内学者也进行了许多有益的探索和讨论。在我国宏观经济预测中有年度预测模型(如国家信息中心的“中国宏观经济模型”(Project LINK)等<sup>②</sup>),但更多的是季度预测模型,如中国人民银行开发的“季度计量经济模型”等<sup>③</sup>。近年来,国内学者对于宏观经济的预测逐渐转向季度预测模型构建并尝试利用高频数据信息,开始采用混频数据模型对我国宏观经济总量进行预测。例如刘汉和刘金全等人(2011)利用社会固定资产投资等指标的月度数据对我国宏观经济总量进行短期预测,认为MIDAS模型在短期预测方面具有精确性的比较优势,在实时预报方面具有可行性和时效性。上述文献在很大程度上丰富和推动我国经济总量预测问题的研究,然而以上文献在预测经济总量时主要是基于最终数据的研究,忽略了数据修正对我国季度GDP预测的影响;在预测因子的选择方面主要使用一致因子,忽略了性质良好的领先因子在季度GDP预测中的表现。

鉴于金融变量作为预测因子的良好性质加之混频数据抽样模型在预测中的明显优势,本文基于混频数据模型在搜集实时数据的基础上,特别关注数据修正对中国季度GDP实时预测的影响,同时重点分析了金融变量(如货币供应量)在混频数据预测中的作用。本文的研究工作具有如下特点与优势:首先,依据相关经济理论本文选择的预测因子具有多样性与新颖性的特征,既包含典型的金融变量如期限结构利差又引入了股票极差波动率这类新型金融变量,它们不但丰富了数据信息量而且与拉动经济增长的“三驾马车”投资、消费、出口密切相关;其次,除利差外我们使用其他变量的同比变化数据,避免季节调整方法选择和环比数据扰动过大对模型估计的影响,同时基于滚动窗口的方法对各类MIDAS模型进行估计,并分析了自回归项在提高模型预测精确性方面的作用;再次,本文使用最终数据和实时数据进行预测来研究数据修正对MIDAS模型估计的影响;最后,我们对组

<sup>①</sup> MIDAS 的全称为 Mixed Frequency Data Sampling.

<sup>②</sup> 该模型为联合国世界计量经济联接模型系统中的中国宏观经济计量模型(1997版)(王慧炯等, 1999)。

<sup>③</sup> 其他中国宏观经济年度与季度模型请参考(王慧炯等, 1999)。

合与多元MIDAS模型进行比较并使用SPA检验方法来考察模型的预测能力。

本文其他部分的结构安排如下：第二部分介绍 MIDAS 模型的设定，第三部分是基于实时数据使用各种 MIDAS 预测模型对 GDP 增长率进行实证分析；最后是本文的主要结论。

## 二、MIDAS预测模型的构建

### (一)、单变量MIDAS( $m, K$ )模型

Ghysels 等人(2004) 首次提出单变量MIDAS( $m, K$ )模型。该模型的鲜明特点是通过含有参数化的多项式权重  $B(L^{1/m}, \theta)$  直接将低频数据  $y_t$  和 高频数据  $x_t^{(m)}$  整合成如下单方程的回归模型。单变量 MIDAS( $m, K$ )模型是最基础的MIDAS模型，其表达形式如下：

$$y_t = \beta_0 + \beta_1 B(L^{1/m}, \theta) x_t^{(m)} + \varepsilon_t \quad (1)$$

其中， $B(L^{1/m}, \theta) = \sum_{k=1}^K w(k, \theta) L^{k-1/m}$ ， $L^{k/m} x_t^{(m)} = x_{t-k/m}^{(m)}$ ； $L^{k/m}$  表示滞后算子可以使得  $x_t^{(m)}$  的值滞后  $k/m$  期； $K$  表示高频数据本身最大滞后阶数； $w(k, \theta)$  是加权函数可以采取不同的形式。

实证研究中加权函数的设定主要有两种形式，一种为指数Almon多项式：

$$w(k, \theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=1}^K \exp(\theta_1 k + \theta_2 k^2)} \quad (2)$$

另一种加权函数为Beta分布多项式：

$$w(k, \theta) = \frac{f(k/K, \theta_1, \theta_2)}{\sum_{k=1}^K f(k/K, \theta_1, \theta_2)} \quad (3)$$

$$f(x, a, b) = \frac{x^{a-1} (1-x)^{b-1} \Gamma(a+b)}{\Gamma(a) \Gamma(b)} \quad (4)$$

其中，

本文对MIDAS类预测模型的估计同时使用包含两个参数的指数Almon多项式和Beta分布多项式作为加权函数，其中指数Almon多项式中的参数一般需要有  $\theta_1 \leq 300$ ， $\theta_2 < 0$  的约束(Clements 和 Galvão, 2008)，而Beta分布多项式的加权函数参数的约束为  $0 \leq \theta_1$ ， $\theta_2 \leq 300$  (Ghysels 等人, 2012)。

### (二)、 $h$ 步向前预测的MIDAS( $m, K, h$ )模型

$h$  步向前预测的MIDAS( $m, K, h$ )模型可表示为：

$$y_t = \beta_0 + \beta_1 B(L^{1/m}, \theta) x_{t-h/m}^{(m)} + \varepsilon_t \quad (5)$$

当  $h=0$  时， $h$  步向前预测MIDAS( $m, K, h$ )模型就退化为MIDAS( $m, K$ )模型。 $h$  的大小变化是高频数据的频率相一致。如果高频数据的频率是月度的， $h$  的大小表示提前的月份。例如：

$$y_t = \beta_0 + \beta_1 B(L^{1/3}, \theta) x_{t-1/3}^{(3)} + \varepsilon_t \quad (6)$$

此时  $h=1$ ，表示使用当前季度提前一个月或者说当前季度中前两个月的信息进行预测。当  $h$  小于 3 时，我们使用当前季度的月度数据进行预测一般称为即时预报(Nowcasting)。

### (三)、 $h$ 步向前预测的ADL-MIDAS( $m, K, h$ )模型

式(1)和式(5)可以被扩展成包含自回归项的MIDAS模型。P阶自回归滞后的ADL-MIDAS( $m, K, h$ )<sup>①</sup>模型可以表示为：

<sup>①</sup> 对于 ADL-MIDAS 模型，本文使用 MIDAS( $m, K, h$ )-AR( $p$ )来表示，这样方便区分自回归项与预测因子的滞后阶数。

$$y_t = \beta_0 + \sum_{j=1}^p \mu_{j+1} y_{t-j} + \beta_1 B(L^{1/m}, \theta) x_{t-h/m}^{(m)} + \varepsilon_t \quad (7)$$

(四)、 $h$ 步向前预测的M-ADL-MIDAS( $m, K, h$ )模型

M-ADL-MIDAS( $m, K, h$ )模型是指多元自回归MIDAS模型。当被解释变量同时受到多个预测因子的影响时可以考虑使用M-ADL-MIDAS( $m, K, h$ )模型。该模型可表示为

$$y_t = \beta_0 + \sum_{j=1}^p \mu_{j+1} y_{t-j} + \sum_{i=1}^s \beta_i B(L^{1/m}, \theta_i) x_{t-h/m}^{i(m)} + \varepsilon_t \quad (8)$$

M-ADL-MIDAS( $m, K, h$ )模型主要缺陷是随着预测因子的增加其所需要估计的参数越来越多，MIDAS( $m, K, h$ )模型简洁性无法保障。因此有些文献关注预测组合(Forecast combination)模型。

(五)、 $h$ 步向前预测的C-ADL-MIDAS( $m, K, h$ )模型

Timmermann (2006)对组合预测方法进行了详细的总结，并认为组合预测可以提高模型的预测精度。C-ADL-MIDAS( $m, K, h$ )模型表示MIDAS组合预测模型，其构造过程如下：

设  $\hat{y}_{i,t+h|t}$  为第 $i$ 个单变量在 $t$ 时刻对  $y_{i,t+h|t}$  样本外预测的结果，组合预测模型表达式如下：

$$f_{t+h|t} = \sum_{i=1}^n w_{i,t} \hat{y}_{i,t+h|t} \quad (9)$$

式(9)中预测权重  $w_{i,t}$  的选择上本文使用比较流行的BMA(Bayesian model averaging)方法<sup>①</sup>。组合预测MIDAS模型中，Sinko 等人(2012)给出的权重表达式为：

$$w_{i,t} = \frac{\exp(-(BIC_i - \min(BIC_i)))}{\sum_{i=1}^n \exp(-(BIC_i - \min(BIC_i)))} \quad (10)$$

其中， $t$ 表示窗口期， $BIC_i$ 表示在第1个窗口期下得到的第 $i$ 个预测结果的BIC信息准则， $n$ 表示预测结果总数。

由于本文将会使用滚动估计方法进行预测，在不同的窗口期下，基于相同变量或者方法的预测结果的BIC信息准则有可能不同。(10)式中的假设忽略了预测时刻 $t$ 的时变性，因此我们对(10)式修改得到时变的权重表达式：

$$w_{i,t} = \frac{\exp(-(BIC_{it} - \min(BIC_{it})))}{\sum_{i=1}^n \exp(-(BIC_{it} - \min(BIC_{it})))} \quad (11)$$

为了衡量MIDAS模型预测的优劣程度，我们需要找到合理的基准预测模型。Clements 和 Galvão(2009)先将高频数据(月度数据)通过算术平均转化低频数据(季度数据)，在同频数据下使用普通最小二乘法得到相应的估计结果，本文亦使用此方法来构造基准预测模型。其表示形式如下<sup>②</sup>：

$$y_t = \beta_0 + \sum_{i=0}^{p-1} \beta_i x_{t-h/m-i} + \varepsilon_t \quad (12)$$

式(12)中  $p = K / m$ ，当  $h = 0$  时等价于同期加总方法获得的同频数据的分布滞后回归模型。模型的预测精度由均方预测误差(MSFE, mean square forecast errors)来度量，预测模型优劣则由均方预测误差的比值即相对均方预测误差(rMSFE) 来表示。

### 三、MIDAS预测模型对GDP增长率的实证分析

<sup>①</sup> BMA 方法的基本思想是使用模型后验分布概率对单变量模型预测结果进行加权，模型后验分布概率又与模型估计得到的 BIC 信息准则相关，即与  $\exp(-BIC/2)$  成比例。

<sup>②</sup> 这里  $m=3$ ,  $x_{t-h/m-i} = (\sum_{s=1}^m x_{t-h/m-i(s)}^{(m)})/m$  相当于使用 3 个高频数据得到 1 个低频数据值。

### (一)、变量选择与混频数据说明

本文选取季度实际GDP同比增长率的实时和最终数据作为低频数据。参照Koenig 等人(2003), 实时数据<sup>①</sup>的选择方法, 我们选取官方初次公布的累计GDP季度数据来计算季度同比增长率作为实时数据, 使用累计GDP季度数据来计算最终数据的实际GDP同比增长率。相关实时数据的收集与处理方法也可具体参照郑挺国和王霞(2010)的研究。样本区间为1992年第一季度至2012年第三季度。

借鉴已有文献, 依据经济金融理论并参照中国经济景气监测中心先行指数中的金融变量, 我们选取股票波动率、货币供应量、期限结构利差、人民币实际有效汇率四个金融变量作为预测因子。金融变量为月度频率的数据在本文中作为高频数据。

根据Romer (1990)的观点股市波动率<sup>②</sup>会通过居民消费支出作用来影响宏观经济总量, 因此股票波动率对宏观经济运行具有一定的前瞻性。本文采用股票波动率作为预测因子并使用已实现极差作为股票波动率的代理变量。我们基于上证综合指数日度数据选取最高价和最低价构造日度极差<sup>③</sup>, 将日度极差累加得到月度已实现极差指标来计算月度同比变化率, 原始数据来源于中国经济数据库。样本区间为1992年9月至2012年9月。

货币供应量是中央银行货币政策中介目标, 同时也为财政政策的实施提供了有力的信贷支持。本文选取货币供应量(M1)的月度同比增长率<sup>④</sup>作为预测因子, 虽然信贷是货币政策传导机制中最重要的变量, 考虑到利率调整及资本市场的影响本文没有选取信贷指标作为预测因子。原始数据来源于锐思金融数据库。样本区间为1994年1月至2012年9月。

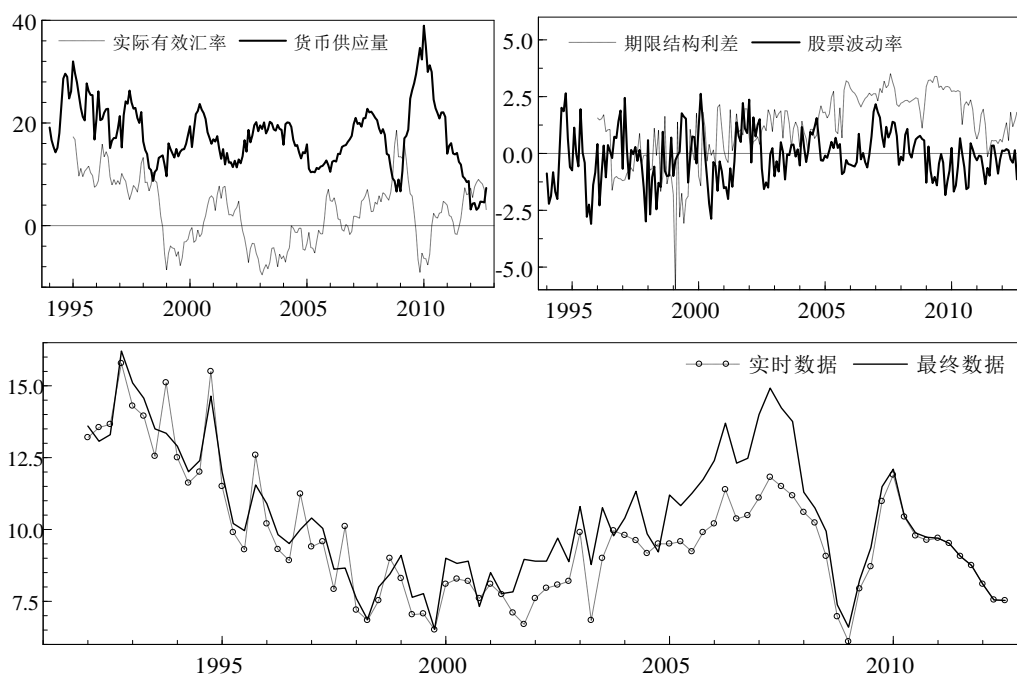


图1 GDP增长率及金融变量混频数据基本情况

期限结构利差是实证文献中经常选用的预测因子。期限结构利差包含远期利率信息其变化会改变人们对未来经济形势的预期。综合期限结构利差的特点及数据的可得性, 具体处理方法是取5年期储蓄存款利率和5至10年期的贷款利率算术平均值作为长期利率, 将算术均值减去3个月同业拆借利

<sup>①</sup>一般的实时数据结构中, 每一列表示特定日期数据 (Vintage data)。Koenig et al.(2003)将特定日期数据区分为: 当前特定日期数据 (current-vintage data)、数终特定日期数据 (end of sample-vintage data)、实时特定日期数据 (real-time-vintage data)。一般所说的实时数据是指数终特定日期数据 (end of sample-vintage data), 它由实时数据结构中的每一列构成, 实时特定日期数据 (real-time-vintage data) 由实时数据结构中的对角线元素构成。

<sup>②</sup>部分学者实证结果表明我国股票市场存在微弱的财富效应, 因此本文没有将股票市场收益率作为预测因子, 由于商品期货市场的数据库较少, 本文亦没有选取商品期货市场变量作为预测因子。

<sup>③</sup>本文使用 Parkinson (1980)的方法来估计极差波动率, 限于篇幅及重要性这里不再陈述具体估计方法。

<sup>④</sup>我们选用货币供应量的代理变量为 M1, 主要因为 M2 的月度数据样本区间较短同时 M2 的数据修订比较频繁。

率的差作为期限结构利差。原始数据来源于中国经济数据库。样本区间为1996年1月至2012年9月。

人民币实际有效汇率是一种贸易加权汇率，它不仅反映汇率资产价格的信息而且包含进出口贸易的基本情况。我们选取人民币实际有效汇率作为预测因子可以反映我国的贸易进出口对宏观经济总量的影响。本文基于人民币实际有效汇率的月度数据得到相应的月度同比变化率。原始数据来源于国际清算银行(IFS)官方网站。样本区间为1995年1月至2012年9月。

本文选取四个金融变量反映了各类金融资产的信息，体现了货币政策目标。作为预测因子不但具有前瞻性、新颖性、多样性的特征而且与拉动经济增长的“三驾马车”即消费、投资、出口密切相关。图1中描述了GDP实时和最终数据的同比增长率以及四个金融变量的混频数据信息。

如前所述，数据修正和预测因子是影响模型预测精度的重要方面，下面将通过实证分析提供经验支持，同时依次分析自回归项、多变量以及组合预测模型在提高混频模型预测精度方面的作用。

## (二)、GDP数据修正的效应分析

本文基于MIDAS模型即混频数据抽样模型进行估计预测，在具体估计时采用滚动窗口方法。我们选择的初始窗口期为1992年7月至2010年1月<sup>①</sup>，样本外预测区间为2010年第1季度至2012年第3季度，共有11个窗口期。另外，我们使用指数Almon和Beta分布多项式两种加权函数进行估计预测，以保证模型估计的稳健性和可靠性。分析数据修正效应之前我们首先给出基本MIDAS模型估计结果。

表1 基于MIDAS(3,K)模型对实时数据估计的MSFE及rMSFE

金融变量	(m, K)	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	rMSFE <sub>beta</sub>	rMSFE <sub>almon</sub>
实际有效汇率	(3,3)	1.1525	1.1525	<b>0.9877</b>	<b>0.9877</b>
	(3,6)	1.1607	1.1671	<b>0.9741</b>	<b>0.9794</b>
	(3,12)	1.0965	1.1427	<b>0.9083</b>	<b>0.9466</b>
	(3,18)	1.3967	1.4417	<b>0.9060</b>	<b>0.9352</b>
期限结构利差	(3,3)	1.0718	1.0727	<b>0.9830</b>	<b>0.9839</b>
	(3,6)	0.9812	0.9641	1.0941	1.0750
	(3,12)	0.6869	0.7299	1.0068	1.0699
	(3,18)	0.7310	0.7200	<b>0.9622</b>	<b>0.9477</b>
股票波动率	(3,3)	1.1481	1.1538	<b>0.9883</b>	<b>0.9932</b>
	(3,6)	1.2289	1.2002	1.0040	<b>0.9805</b>
	(3,12)	1.3474	1.3196	1.0079	<b>0.9871</b>
	(3,18)	1.4231	1.2686	<b>0.9983</b>	<b>0.8899</b>
货币供应量	(3,3)	0.7069	0.7204	1.1737	1.1962
	(3,6)	0.6798	0.6695	1.0928	1.0762
	(3,12)	0.6468	0.5969	1.2207	1.1265
	(3,18)	0.5430	0.5999	<b>0.6750</b>	<b>0.7457</b>

注：表中m表示季度数据与月度数据的关系其取值为3，K表示预测因子的最大滞后阶数，MSFE<sub>beta</sub>和MSFE<sub>almon</sub>分别表示Beta分布和指数Almon多项式下得到的MSFE。

表1中汇总了基于MIDAS(3,K)模型得到的MSFE和rMSFE估计结果。我们这里给出的滞后阶数有3、6、12、18四种情况<sup>②</sup>。首先，从表中我们可以发现除股票波动率外，增加滞后阶数K可以减少相应的MSFE。实际有效汇率和期限结构利差的最优滞后阶数均为12，而货币供应量的最优滞后阶数为18。其次，各个变量对GDP增长率的预测精度存在差异，货币供应量与期限结构利差的预测精度

<sup>①</sup> 滚动窗口是窗宽设置的太小会损失样本量。鉴于本文样本量有限，我们设置窗宽较大。由于每个变量的样本量不同，初始窗宽实际上是不同的，初始窗口期只是一个窗口范围，可以通过程序设计保证每个变量在此范围选择合适的窗宽。没有特别说明本文其他类型的MIDAS模型滚动窗口个数不变。

<sup>②</sup> Clements and Galvão (2009)认为对与季度GDP数据而言，基于月度数据的预测因子最大滞后阶数K要选择3的倍数以方便与基准模型进行比较在实证研究中最大滞后阶数K最长可以选到2年，鉴于中国的数据较短本文主要考虑1季度、半年、1年、一年半四种情况，分别对应的滞后阶数K为3、6、12、18。

较好，股票波动率的预测精度不够理想。最后，从 $rMSFE$ 估计结果来看，MIDAS(3,K)模型总体优于基准预测模型，说明混频模型在预测方面具有比较优势。

以上分析完全基于实时数据采用MIDAS模型进行估计，数据修正是否对MIDAS模型的预测结果产生负面影响是本文关注的重要问题之一。如果数据修正对模型估计产生负面影响，则说明基于实时数据的预测会更加合理。因此，本文使用单变量MIDAS模型分别对GDP增长率的实时数据和最终数据进行估计，其相应的估计结果汇总在图2中。

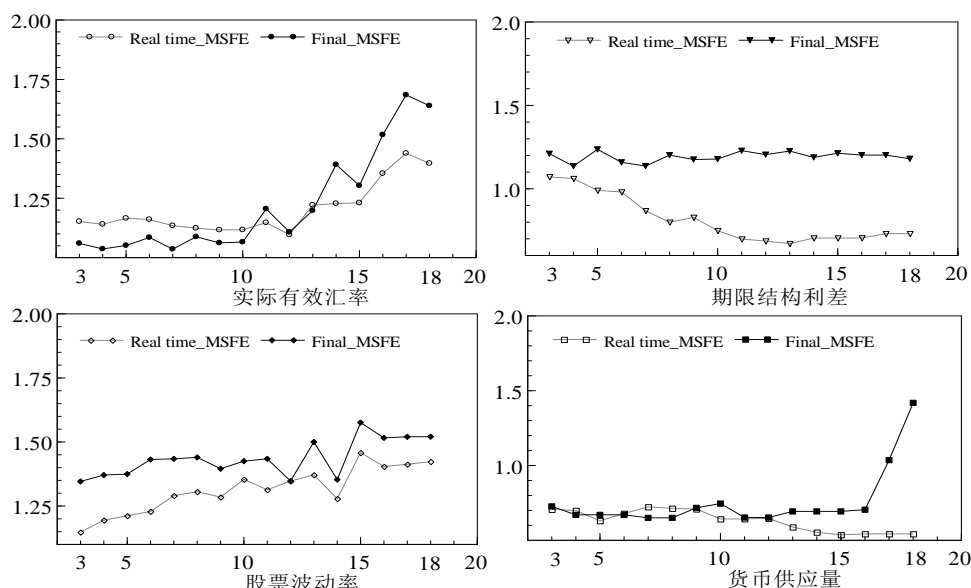


图2 实时数据与最终数据估计的MSFE对比图

图2汇总了单变量MIDAS模型下实时和最终数据MSFE的估计情况，其中预测因子的最大滞后阶数K从3取至18并选取Beta分布加权多项式作为加权函数。从图中我们可以看出随着最大滞后阶数的增加，基于最终数据估计的MSFE大于实时数据的相应估计结果，说明数据修正对MIDAS模型的预测结果产生额外的“噪音”。特别地，当滞后阶数越大即包含信息越多时，数据修正的负面影响也会越大。本文认为采用实时数据可以减少数据修正的负面影响，进而可以提高MIDAS模型的预测精度。该结果与Clements and Galvão (2009)的研究结论一致，有鉴于此下文的分析全部基于实时数据。

### (三)、金融指标的预测能力分析

如前所述，我们认为预测因子对模型预测精度有重要影响，为此本节比较分析了金融指标与其他宏观经济指标的预测效果。而对于宏观经济变量如GDP增长率来说，自回归项包含在解释变量中会使得模型的预测效果更好(Clements 和 Galvão, 2009)。在比较分析之前我们首先关注了在给定预测因子不变的情况下，MIDAS( $m, K, h$ )-AR( $p$ )模型在改进预测精度方面所起到的作用，其次基于该模型对金融指标与其他宏观经济指标的预测效果进行比较分析。

在分析包含自回归项的MIDAS模型时。本文使用AIC信息准则来确定最优的自回归项滞后阶数 $p$ 。考虑自回归项滞后0阶至4阶的情况，从中选取最小的AIC信息准则值相对应的滞后阶数作为最优滞后阶数。实证结果表明对于所有变量来说，最小的AIC信息准则值对应的滞后阶数均不为0，说明在解释变量中包含自回归项的模型设定更加合理。

表2汇总了MIDAS(3,K,0)-AR( $p$ )模型的MSFE及 $rMSFE$ 估计结果。首先，相对应与表1中的估计结果来说，表2中MSFE的值大幅下降，意味着包含自回归项MIDAS模型预测精度相比基本MIDAS模型大幅提高<sup>①</sup>。说明自回归项在改进MIDAS模型预测精度方面发挥了重要作用。其次，我们还可以发现在最优滞后阶数下，期限结构利差和货币供应量的MSFE取值较小，实际有效汇率次之。结合前文的分析得知期限结构利差和货币供应量对拉动经济增长的投资因素有重要影响，而实际有效汇

<sup>①</sup> 与表1相比加入自回归项之后，货币供应量相对应的MSFE的值没有太大变化。

率与出口因素密切相关,这说明投资和出口仍然是我国促进经济增长的主要手段。同时与消费因素相对应的股票波动率的预测精度相对较差,反映了我国内需不足的现实。该结果与刘汉和刘金全等人(2011)的研究结论一致。

表2 MIDAS(3,K,0)-AR(p)<sup>①</sup>模型的实时数据 MSFE 及 rMSFE 估计结果

金融变量	(m,K,p)	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	rMSFE <sub>beta</sub>	rMSFE <sub>almon</sub>
实际有效汇率	(3,3,4)	0.8855	0.8894	<b>0.9959</b>	1.0002
	(3,6,4)	0.8197	0.8602	<b>0.9281</b>	<b>0.9740</b>
	(3,12,4)	0.8026	0.8333	<b>0.8697</b>	<b>0.9030</b>
	(3,18,1)	0.7123	0.7486	<b>0.7141</b>	<b>0.7504</b>
期限结构利差	(3,3,1)	0.6435	0.6433	1.0471	1.0467
	(3,6,1)	0.6078	0.6035	1.0501	1.0426
	(3,12,1)	0.5466	0.5548	<b>0.9794</b>	<b>0.9941</b>
	(3,18,1)	0.5674	0.5356	<b>0.9692</b>	<b>0.9149</b>
股票波动率	(3,3,4)	1.0289	1.0831	<b>0.9584</b>	1.0089
	(3,6,4)	1.0825	1.0767	1.0016	<b>0.9962</b>
	(3,12,1)	1.0807	1.0767	1.0045	1.0162
	(3,18,1)	0.9434	0.9394	<b>0.9978</b>	<b>0.9936</b>
货币供应量	(3,3,4)	0.5640	0.5692	<b>0.9190</b>	<b>0.9275</b>
	(3,6,4)	0.5574	0.5575	<b>0.8476</b>	<b>0.8477</b>
	(3,12,4)	0.7646	0.5664	<b>0.8999</b>	<b>0.6665</b>
	(3,18,4)	0.8308	0.5955	1.0637	<b>0.7625</b>

另外,我们还对  $h$  步向前预测MIDAS( $m,K,h$ )-AR( $p$ )模型进行估计,其中  $h$  取值为1至12,我们发现MIDAS-AR( $p$ )模型更加适用于短期预测。限于篇幅及重要性这里没有给出具体估计结果。

下面基于MIDAS( $m,K,h$ )-AR( $p$ )模型对金融变量的预测效果与投资,消费,出口相应的预测效果进行比较。参照刘汉和刘金全(2011)变量选取方法,我们选取固定资产总额增速,社会零售总额增速,进出口增速分别反映投资、消费和出口,同时根据前文的分析将金融变量分别对应投资、消费和出口以方便比较。从表3报告的即期预报( $h=0$ )的估计结果来看,与投资 and 出口相对应的金融指标预测效果优于投资和出口指标,而波动率指标的预测表现略逊于消费指标。但是从短期预测( $h=1$ )的结果来看,与投资、消费、出口相对应的金融指标的预测效果全部占优,而且消费指标的预测效果明显不如投资和出口指标。这表明金融变量在预测中体现了领先因子的性质,我们认为选取包含更多未来信息的金融变量对季度GDP进行预测更为合理。

表3 MIDAS( $m,K,h$ )-AR( $p$ )模型下金融指标与投资、消费、出口的预测结果比较

	h=0							
	K=3		K=6		K=12		K=18	
	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>
投资	1.1071	1.1071	1.0700	1.0532	1.0321	0.9368	0.9545	0.9291
货币供应量	0.5640	0.5692	0.5574	0.5575	0.5664	0.7646	0.8308	0.5955
期限结构利差	0.6435	0.6433	0.6078	0.6035	0.5466	0.5548	0.5674	0.5356
消费	1.0191	1.0191	1.0726	1.0679	0.9538	0.8442	0.9287	1.0198
股票波动率	1.0289	1.0831	1.0825	1.0767	1.0933	1.0807	0.9434	0.9394

<sup>①</sup> 由于本文使用的滚动窗口估计方法,在对每一个窗口期自回归项的最优滞后阶数有可能不同,表3(m,K,p)中  $m$  表示月度频率与季度频率的关系始终为3;  $K$  表示高频金融变量选取的最大滞后阶数这里选取3,6,12,18四组;  $p$  的取值表示所有窗口期中取得次数最多的自回归项的最优滞后阶数。如实际有效汇率,当滞后3阶时有8个窗口期取得的最优滞后阶数为4,有一个窗口期取得阶数为2,有两个窗口期取得的滞后阶数为1,最终在表3中表示为(3,3,4)。



出 口	1.0487	1.0385	1.0451	1.0339	0.9636	0.9906	1.1239	0.9772
实际有效汇率	0.8855	0.8894	0.8197	0.8602	0.8026	0.8333	0.7123	0.7486
h=1								
	K=3		K=6		K=12		K=18	
	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>	MSFE <sub>beta</sub>	MSFE <sub>almon</sub>
投 资	1.1214	1.1156	1.1278	1.0932	1.0523	1.0568	0.9926	0.9980
货币供应量	0.5574	0.5574	0.6469	0.5650	0.6839	0.6509	1.0110	0.5808
期限结构利差	0.6036	0.5914	0.5681	0.5668	0.5249	0.5406	0.5308	0.5511
消 费	1.3561	1.3561	1.5280	1.4726	1.3338	1.2578	1.2904	1.2828
股票波动率	1.0567	1.0595	1.0479	1.0641	1.0263	0.9300	0.9881	0.8714
出 口	1.0932	1.0932	1.1360	1.0955	1.0068	1.0072	1.1062	0.9975
实际有效汇率	0.8875	0.8991	0.8857	0.8935	0.7871	0.8050	0.7843	0.7993

表5 组合MIDAS(3,K,h)-AR(p)与多元MIDAS(3,K,h)-AR(p)的MSFE估计结果

			h=0	h=1	h=2	h=6	h=9	h=12
K=3	C-MIDAS	F_BIC	0.6434	0.5994	0.5944	<b>0.9046</b>	1.0045	0.8269
		D_BIC	0.6851	0.6028	0.5947	<b>0.9046</b>	1.0045	0.8269
	M-MIDAS	Almon	0.7166	0.7554	0.7487	<b>0.8385</b>	1.0965	0.9899
		Beta	0.7924	0.7554	0.7706	<b>0.8385</b>	1.1053	1.0343
K=6	C-MIDAS	F_BIC	0.6051	0.5668	0.5608	<b>0.8885</b>	0.9630	0.9561
		D_BIC	0.6060	0.5682	0.5606	<b>0.8885</b>	0.9633	0.9561
	M-MIDAS	Almon	0.6851	0.7247	0.7919	<b>0.8410</b>	0.8277	1.3472
		Beta	0.7026	0.7424	0.7909	<b>0.7937</b>	1.0670	1.5085
K=12	C-MIDAS	F_BIC	0.5506	0.5404	0.5422	0.8937	0.8090	0.9489
		D_BIC	0.5535	0.5406	0.5420	0.8928	0.8100	0.9522
	M-MIDAS	Almon	0.7974	0.7134	0.6045	0.9744	1.1302	1.5883
		Beta	0.6599	0.5025	0.6311	1.0863	1.3015	1.6384
K=18	C-MIDAS	F_BIC	0.5364	<b>0.5406</b>	0.5468	0.8566	0.6450	0.9741
		D_BIC	0.5315	<b>0.5465</b>	0.5462	0.8523	0.6563	0.9831
	M-MIDAS	Almon	0.5118	<b>0.5001</b>	0.5968	0.9287	0.9421	1.6339
		Beta	0.6215	<b>0.5113</b>	0.5485	1.2600	1.4622	1.6658

注：表中K表示最大滞后阶数，h表示预测步长；C-MIDAS表示组合MIDAS(3,K,h)-AR(p)模型，M-MIDAS表示多元MIDAS(3,K,h)-AR(p)；F\_BIC和D\_BIC分别表示不同权重下的组合预测，其中F\_BIC与文中式(10)对应，D\_BIC则与式(11)对应；Almon和Beta分别表示加权函数基于指数Almon和Beta分布多项式的M-MIDAS模型。

为了说明组合MIDAS-AR模型预测能力占优于多元MIDAS-AR模型，本文借鉴Hansen（2005）提出的模型预测能力检验方法即SPA（Superior Predictive Ability）检验方法。该方法原假设为基准或目标预测模型不差于其他预测模型，并通过损失函数定义相对损失变量，基于相对损失变量构造检验统计量，限于篇幅及重要性这里我们没有给出统计量数学表达式。该统计量在原假设下基于Bootstrap方法得到SPA检验相应的P值。我们用于检验的目标预测模型为组合MIDAS-AR模型，其他预测模型主要包括不同加权函数下的多元MIDAS-AR模型。表6报告了基于10000次Bootstrap抽样得到SPA检验估计结果。从表中可以看出几乎所有的SPA检验P值都很难拒绝原假设，组合MIDAS-AR模型预测能力不差于多元MIDAS-AR模型，同时上文中证实结果表明组合MIDAS-AR模型预测精度更好，因此，我们认为组合MIDAS-AR模型预测能力占优。

表6 模型预测能力SPA检验结果汇总(P-Values)

h	K=3		K=6		K=12		K=18	
	DS	DA	DS	DA	DS	DA	DS	DA
0	0.5271	0.1741	0.5861	0.3049	0.4791	0.5321	0.6275	0.6238
1	0.5271	0.5575	0.5947	0.5037	0.2706	0.1591	0.5075	0.4354
2	0.5067	0.5846	0.0888	0.5651	0.4992	0.5541	0.6638	0.2635
6	0.1681	0.1541	0.1541	0.1537	0.5902	0.4358	0.0444	0.4664
9	0.4923	0.2939	0.2187	0.2285	0.5584	0.5059	0.0009	0.0052

注：K表示滞后阶数，h表示预测步长，DS和DA表示构造统计量的相对损失变量分别方差形式和绝对值形式。

## 四、主要结论

对宏观经济总量进行合理预测,准确把握经济增长运行态势是当前宏观经济与政策研究的重大课题之一。本文综合考察了可能影响宏观经济总量预测效果的三个重要因素,即数据修正、预测因子、数据频率。基于实时数据通过MIDAS模型来分析数据修正的影响并考察了混频金融变量在季度GDP即期预报和短期预测中的作用。我们依次分析了自回归项,多变量以及组合预测模型在提高混频数据抽样模型预测精度方面的作用,并使用SPA检验方法衡量模型的预测能力。

根据前面的实证分析,本文主要结论可概括如下四个方面:一是MIDAS模型总体上好于基准模型而且短期预测效果甚佳,若在模型中包含自回归项则可以明显提高预测精度,在各类MIDAS模型中组合MIDAS-AR模型的预测精度更好,预测能力占优;二是基于MIDAS-AR模型可以发现,货币供应量、期限结构利差和实际有效汇率的短期预测精度占优,进一步结果表明投资和出口依旧是拉动我国经济增长的两个重要因素;三是通过MIDAS模型对最终数据和实时数据进行对比分析可以发现数据修正对于预测精度有负面影响;四是通过比较金融指标与经济变量的预测效果我们发现金融变量具有较好的预测能力,体现了领先因子的作用。

总之,金融指标在季度GDP实时预报和短期预测发挥了重要作用,不但拥有明显的时效性而且具有较好的预测效果;对GDP进行预测时实时数据可以明显提高MIDAS模型的预测精度。因此使用预测精度更好、预测能力占优的MIDAS模型进行实时预测能够体现出一定时效性和精确度,这对准确把握我国经济增长运行态势将具有重要的参考价值。然而,我国宏观经济总量的经济周期性质还没有在MIDAS模型中充分反映出来,因此如何在MIDAS模型中融入经济周期因素对季度GDP进行短期预测是我们需要进一步研究的问题。

## 参 考 文 献

- [1] 刘汉和刘金全, 2011,《中国宏观经济总量的实时预报与短期预测—基于混频数据预测模型的实证研究》,《经济研究》第3期4-17页。
- [2] 刘金全和刘兆波, 2008,《我国货币政策的中介目标与宏观经济波动的关联性》,《金融研究》第10期37-47页。
- [3] 刘金全、刘汉和印重, 2010,《中国宏观经济混频数据模型应用—基于MIDAS模型的实证研究》,《经济科学》第5期23-34页。
- [4] 王慧炯、李泊溪和李善同, 1999,《中国实用宏观经济模型1999》,中国财政经济出版社。
- [5] 郑挺国、王霞, 2013,《中国经济周期的混频数据测度及实时分析》,《经济研究》即将发表。
- [6] 郑挺国、王霞, 2010,《中国产出缺口的实时估计及其可靠性研究》,《经济研究》第10期129-142页。
- [7] Bansal, R. and A. Yaron 2004. "Risks for the long run: A potential resolution of asset pricing puzzles." *The Journal of Finance* 59(4): 1481-1509.
- [8] Burkhard Raunig and Johann Scharler, 2010, "Stock Market Volatility and the Business Cycle", *Monetary Policy & The Economy*, Q2, pp.54-63.
- [9] Camacho M., and G. Perez-Quiros, 2010, "Introducing the EURO-STING: short-term indicator of Euro area growth", *Journal of Applied Econometrics*, 25, pp.663-694.
- [10] Clements, M. P. and Galvão A. B., 2008, "Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting US Output Growth", *Journal of Business and Economic Statistics*, 26 (4), pp.546—554.
- [11] Clements, M. P. and Galvão A. B., 2009, "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models", *Journal of Applied Econometrics*, 24(7), pp.1187—1206.
- [12] Clive W.J. Granger and Allan Timmermann, 2006, "Forecast Combinations," forthcoming in *Handbook of Economic Forecasting*, edited by Graham Elliott.
- [13] Elena Andreou, Eric Ghysels and Andros Kourtellis, 2010, "Regression models with mixed sampling frequencies.", *Journal of Econometrics*, 158, pp.246-261.
- [14] Evan F. Koenig, Sheila Dolmas, and Jeremy Piger, 2003, "The use and abuse of real time data in economic forecasting",

The Review of Economics and Statistics, 85(3), pp. 618–628.

- [15] Galv ã A. B, 2010, “The role of high frequency data and regime changes in predicting economic activity with financial variable” ,working paper.
- [16] Ghysels, E.,Santa-Clara, P. and Valkanov R.,2004, “The MIDAS Touch: Mixed Data Sampling Regressions” , mimeo, Chapel Hill, N. C.
- [17] Ghysels, E. and Wright, J. H.,2009, “Forecasting professional forecasters”, Journal of Business and Economic Statistics, 27(4), pp.504-516.
- [18] Ghysels, E .Arthur Sinko and Rossen Valkanov ,2007,“MIDAS Regressions: Further Results and New Directions”, Econometric Reviews, 26(1) ,pp.53-90.
- [19] Guerin P and Massimiliano Marcellino,2011, “Markov-switching MIDAS models.” EUI working paper.
- [20] Harvey, A. ,1989, “Forecasting, Structural Time Series Models and the Kalman Filter”, Cambridge University Press, Cambridge.
- [21] Heij C, D.v.Dijk and P.J.F. Groggen,2011,“Real time macroeconomic forecasting with leading indicators: An empirical comparison.”, International Journal of Forecasting, 27, pp.466-481.
- [22] Libero Monteforte and Gianluca Moretti, 2010, “Real time forecasts of inflation: the role of financial variables”, working paper, No. 767.
- [23] JD Hamilton and G Lin,1996, “Stock market volatility and the business cycle” , Journal of Applied Econometrics, 11(5) , pp. 573–593.
- [24] Parkinson, M., 1980, “ The extreme value method for estimating the variance of the rate of return”, Journal of Business 53(1), pp.61–65.
- [25] PR Hansen,2005, “A test for superior predictive ability.”,Journal of Business & Economic Statistics, 23(4) pp. 365-380.
- [26] Romer, C, 1990, “The Great Crash and the Onset of the Great Depression ”,The Quarterly Journal of Economics,105, pp.597–624.
- [27] Sinko.A , Sockin .M and Ghysels, 2012, Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models. Chapel Hill, N. C.
- [28] Stock, J. H. and Watson, M. W.,2003, “How Did Leading Indicator Forecasts Perform During the 2001 Recession? ” Federal Reserve Bank of Richmond: Economic Quaterly,89 (3), pp. 71–90.

## Mixed-Frequency Based Forecasting China’s GDP with Financial Indices

Zheng Tingguo      Shang Yuhuang

(The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian, 361005)

**Abstract:** The paper employs mixed frequency data sampling (MIDAS) model to nowcast and forecast GDP with financial variables in real time data. We find that: The Prediction of MIDAS model is very well in short-term and the model with autoregressive term significantly reduce the forecast error; Data revision have a negative impact on accuracy of prediction with the MIDAS model; Money supply and other predictors have the higher accuracy of prediction in MIDAS-AR model, which means investment and exports are still two important factors driving China's economic growth; SPA test prove combination of MIDAS model is no lass then other model and can better improve the accuracy of prediction that means the prediction power of combination of MIDAS model is dominant.

**Key words:** Financial Variable; GDP growth rate; MIDAS model; Mixed-frequency forecasting; Data revision